

BASiD: Biografiedaten ausgewählter Sozialversicherungsträger in Deutschland - Erstellung eines Längsschnitt-Biografiedatensatzes auf Basis der Daten der Deutschen Rentenversicherung, der Bundesagentur für Arbeit und des Instituts für Arbeitsmarkt- und Berufsforschung

Ziel des Projekts:

In diesem Projekt werden erstmalig Personendaten über die Grenzen zweier Daten haltenden Institutionen zu einem Biografiedatensatz zusammen gespielt. Große Teile der Personendaten der Bundesagentur für Arbeit (BA) und des Instituts für Arbeitsmarkt- und Berufsforschung (IAB) sowie der deutschen Rentenversicherung (RV) stammen aus dem gleichen Datengenerierungsprozess und bauen auf der gleichen rechtlichen Grundlage, dem Sozialgesetzbuch auf. Ziel dieses Projektes ist es einen gemeinsamen Datensatz aus den Daten der RV und der BA bzw. des IAB zu erstellen und der Wissenschaft als Scientific Use File (SUF) und in schwach anonymisierter Form im Gastaufenthalt zur Verfügung zu stellen.

Generierung des Datensatzes/Ablauf des Projekts:

Die Versicherungskontenstichprobe (VSKT)¹ der Rentenversicherung ist der Basisdatensatz für dieses Projekt. Alle Personen, die sich in dieser Stichprobe befinden, werden entsprechend in den Datenbanken des IAB ermittelt. Die Datenquellen sind hierbei die Beschäftigten- und Leistungsempfänger-Historik (BLH), die Integrierten Erwerbsbiographien (IEB) und das Betriebs-Historik-Panel (BHP).² Durch das Matching von Informationen dieser Datenquellen wird erreicht, dass für die von der Sozialversicherung erfassten Personen im Alter von 15 bis 67 Jahren, empirische Informationen vom ersten Beitrag bis zum aktuellen Rand vorliegen. In der anfänglichen Projektphase werden die Variablen, die in beiden Datensätzen vorliegen, zusammengespielt und geprüft. Schwierigkeiten ergeben sich hierbei aus der Tatsache, dass Informationen nicht in der gleichen Form bei beiden Institutionen vorliegen bzw. parallele Zustände im Datensatz auftauchen können. Beispielsweise weicht im Bezug auf die Anzahl der Arbeitslosen die offizielle Statistik der BA von der der RV ab, da diese bei beiden Institutionen unterschiedlich erfasst wird. Um mit den Schwierigkeiten adäquat umgehen zu können müssen Heuristiken entwickelt werden, weshalb die Datenprüfung bzw. -bereinigung mit einem erheblichen Zeitaufwand verbunden ist.

Nach Abschluss der Datenbereinigung und ersten Zusammenfassungen von Ausprägungen aus Datenschutzgründen, werden die über die gemeinsamen Variablen hinausgehenden zusätzlichen Merkmale aus beiden Datenquellen hinzugespielt. Schließlich wird dem Datenschutz ein Vorschlag für einen SUF unterbreitet. Der so entstandene Datensatz soll in einem weiteren Schritt zur Testnutzung angeboten werden, um eventuelle Verbesserungsmöglichkeiten in die endgültige Version des SUF einarbeiten zu können. Da Forschungsfragen auf regionaler Basis immer stärker an Bedeutung gewinnen, wird im Rahmen des Projektes zusätzlich eine Machbarkeitsstudie für die Erstellung eines SUF für Regionalanalysen durchgeführt. Da auch für ausländische Wissenschaftler die Nutzung des

¹ Die VSKT ist eine ständige Sondererhebung der gesetzlichen Rentenversicherung. Sie erfasst die Versicherten der gesetzlichen Rentenversicherung und den Stand ihrer Rentenanwartschaften in Deutschland und liefert Informationen über sämtliche (gespeicherten) rentenrelevanten Tatbestände, das heißt Informationen zu rentenrechtlichen Zeiten und Beiträgen, die nach einer Kontenklärung im Rentenversicherungskonto der Person gespeichert sind.

² Die BLH umfasst für den Zeitraum ab 1975 eine lückenlose Historie zu den sozialversicherungspflichtig Beschäftigten und Personen mit Leistungsbezug. Die IEB enthält ebenfalls Informationen zu den sozialversicherungspflichtig Beschäftigten und Personen mit Leistungsbezug sowie Informationen zu Zeiten der Arbeitssuche und Maßnahmeteilnahmen. Das BHP umfasst alle Betriebe in Deutschland mit mindestens einem sozialversicherungspflichtig Beschäftigten zum 30.6. eines Jahres. Ab 1999 sind zusätzlich Betriebe enthalten, die keine sozialversicherungspflichtig Beschäftigten, aber mindestens einen geringfügig Beschäftigten aufweisen.

SUF gewährleistet werden soll, werden in der Projektzeit Dokumentationen sowohl in Deutsch als auch in Englisch erstellt.

Inhalte des Biographiedatensatzes:

Der neue Datensatz wird sich durch verschiedene Charakteristika von den bisherigen Datensätzen abheben: Er beinhaltet zum einen mehr Merkmale, die es ermöglichen Fragestellungen zu bearbeiten, die bisher für Deutschland nicht bzw. nur sehr ungenau beantwortet werden konnten. Zweitens bildet der Datensatz die komplette Biografie der Untersuchungsperson auch vor 1975 ab und ist somit für keine Person links zensiert.³

Neben den datentechnischen und demographischen/regionalen Merkmalen, finden sich Variablen der RV, wie zum Beispiel Rentenzeiten, Anrechnungszeiten, Berücksichtigungszeiten, Entgeltpunkte und Geburtsangaben zu Kindern der im Datensatz enthaltenen Personen im SUF. Die zugespielten Daten der BA bzw. des IAB sind Betriebsinformationen, wie die Betriebsgröße, der Wirtschaftszweig und Informationen zu den Beschäftigten, sowie Daten aus dem Meldeverfahren der Sozialversicherung. Diese beinhalten zum Beispiel Informationen zu Ausbildungs- Beschäftigungs- und Arbeitslosigkeitszeiten, Zeiten der Arbeitssuche, Teilnahme an Maßnahmen, sowie Informationen zum Entgelt.

Somit lassen sich u.a. Analysen:

- zu Geburten und Erwerbsverläufen von Frauen,
- zu Aussiedlern,
- zu Krankheit, Pflege im Erwerbsverlauf,
- zum Einfluss von Militärdienst oder Zivildienst auf den Erwerbsverlauf,
- zum Lebenseinkommen und Entgeltpunkten für die Rentenberechnung,
- zum Einfluss von Startbedingungen im Arbeitsmarkt auf den Karriereverlauf,
- zu Änderungen von gesetzlichen Regelungen um den Einführungszeitpunkt und
- Analysen auch für Regionen/Bundesländer

durchführen.

Beide Forschungsdatenzentren wollen mit dem neu entstehenden Datensatz im Laufe des Projektes eigene Forschungsthemen bearbeiten wie. z.B. eine Machbarkeitsstudie für die Erstellung eines SUF für Regionalanalysen, um die Vor- und Nachteile des Datensatzes zu ermitteln und Verbesserungen vorzunehmen.

Datenangebot und -zugang:

Der Wissenschaft und Forschung im In- und Ausland soll dieser Datensatz kostenlos in Form eines SUF, der über beide Forschungsdatenzentren bezogen werden kann, zur Verfügung gestellt werden. Die Daten werden hauptsächlich als SUF angeboten, da der Zugang über einen SUF immer noch der bevorzugte Zugang zu einem Datensatz von Seiten der Wissenschaft ist und der SUF ins Ausland verschickt werden kann. Darüber hinaus wird der schwach anonymisierte Datensatz im Gastaufenthalt angeboten werden, um detaillierte Analysen (z.B. für Regionen oder Maßnahmentypen) zu ermöglichen.

Information:

Informationen zum Projekt werden auf den Internetseiten des FDZ der BA/IAB unter http://fdz.iab.de/de/FDZ_Projects.aspx bekannt gegeben.

³ Dies gilt natürlich schon für die angebotene Datensätze der RV, nicht jedoch für die Datensätze des Forschungsdatenzentrums der BA im IAB, die frühestens 1975 einsetzen.